International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 6, No. 1, 2024

Bioinformatics Analysis of Core Genes Related to Sarcopenia

Qing Lin [1, a], Fengxu Wei [2, *]

[1] Baise center for disease control and prevention, Baise, Guangxi, China
[2] Youjiang medical university for nationalities, Baise, Guangxi, China

*** Corresponding author:** Fengxu Wei (Email: 1449318783@qq.com), [a] 1751250936@qq.com

Abstract: Objective: To analyze the genes related to the pathogenesis of sarcopenia (SA) and find the potential therapeutic targets. Methods: GSE1428 data set downloaded from GEO gene expression omnibus, GEO) was taken as the analysis object. The data were standardized by R software, and the differential genes were screened. The GO and KEGG function enrichment analysis of the differential genes was carried out by Metascape, and the PPI analysis was carried out by String. The results were imported into Cytoscape software to screen out the core genes, which were imported into BioGPS database for tissue localization. Results: 117 differential genes were obtained, including 54 down-regulated genes and 63 up-regulated genes. GO and KEGG analysis showed that GO was mainly enriched in ethanol reaction, carboxylic acid transport, cell secretion regulation, vitamin metabolism, leukocyte chemotaxis regulation and bone marrow cell differentiation regulation, while KEGG was mainly enriched in viral carcinogenesis pathway. There are 108 nodes and 44 connections in the PPI network. According to the PPI results, Cytoscape screened out core genes such as RPE65, TOPORS, IMPG2 and IMPDH1, among which RPE65, TOPORS and IMPDH1 were down-regulated and IMPG2 was up-regulated. The localization of gene tissue suggested that RPE65 was located in retina. Conclusion: Screening the core genes and related signal pathways is helpful to understand the pathogenesis of SA and provide new ideas for drug therapy research of SA.

Keywords: Sarcopenia; Nosogenesis; Core Gene; Bioinformatics.

1. Background

Myopenia is a disease related to aging and characterized by the progressive decline of skeletal muscle mass, strength and performance[1],The elderly are frequent, which affects the quality of life in the later period. According to epidemiological investigation, the number of patients with SA in the world has reached 50 million at present, and it is expected to reach 500 million in 2050[2]; The prevalence rate of SA in China is 5.5% ~ 25.7%, among which the prevalence rate of people over 60 years old is 9.3%, 6.4% for men and 11.5% for women[3]; The global prevalence rate of SA is about 13% ~ 24% among people aged 65-70, and the prevalence rate of people over 80 is as high as 50%[4].Muscle and bone are derived from the same mesenchymal stem cells[5],SA and osteoporosis (OP) often appear together, and SA is often found after brittle fracture. The incidence of brittle fracture and fall in patients with SA increases, and the quality of life in the later stage decreases obviously[6].Therefore, timely diagnosis and treatment of SA is particularly important to improve the quality of life in old age.

The onset of SA involves aging, inflammation, hormonal imbalance and mitochondrial dysfunction[7-9],At present, the research progress of SA is limited by the lack of unified diagnostic criteria, the difficulty in obtaining muscle biopsy samples and the lack of animal or cell models similar to human body. With the development of genome-wide association studies and the second generation sequencing technology, the number of biological databases has obviously increased, and using bioinformatics to find related pathogenic genes has become a hot research topic. Therefore, this study is based on GEO public database, and the data set GSE1428 is taken as the analysis object. By screening differential genes →GO and KEGG enrichment analysis → constructing protein interaction network → screening core genes → locating gene organization, the biological function of SA core genes is deeply understood, so as to screen out the key pathogenic factors of SA.

2. Data and Methods

2.1. Information

This study is based on GEO public database (https://www.ncbi.nlm.nih.gov/geo/) for data retrieval. The data set GSE1428 is the research object, and the gene chip information platform is GPL96, including 10 young people and 12 elderly SA samples, all of which are taken from the lateral thigh muscle. The inclusion criteria for young people are 19 to 25 years old, and the inclusion criteria for the elderly are 70 to 80 years old.

2.2. Method

2.2.1. Differential Gene Expression Analysis

In this study, limma package in R language was used to analyze the differential expression of SA's mRNA chip raw data, and the screening conditions of test statistics were set as $P<0.05$ and the logarithmic absolute value of difference multiple (FC) $|\log FC|>1$ to screen Differentially expressed gene (DEGs) in this data set.

2.2.2. Enrichment Analysis of GO and KEGG

DEGs obtained from the above screening were made into Excel tables and imported into Metascape online analysis tools for differential gene enrichment analysis.

2.2.3. PPI and Core Gene Module Analysis

DEGs was analyzed by STRING tool, and the protein-protein interaction network (PPI) was obtained. The calculation results of STRING were analyzed by Cytoscape 3.9.1 software, and the core genes of SA were found by using

plug-ins MCODE and Cytohubba.

2.2.4. Core Gene Mapping

The core gene was located by BioGPS online database, and the highest expression between tissues was more than twice the second expression, which indicated that the gene could be used as a marker gene for SA expression.

3. Result

3.1. Differential Gene Acquisition

GSE1428 is analyzed by R software, and the next step is analyzed under the condition that the gene expression of database samples is basically the same, as shown in Figure 1. Then, set the test statistics P<0.05 and the logarithmic absolute value of FC |log FC|>1 as the screening conditions, and use the limma package of R language to screen the differential genes, and get 117 DEGs, 54 down-regulated genes and 63 up-regulated genes, and draw a volcano map, as shown in Figure 2.

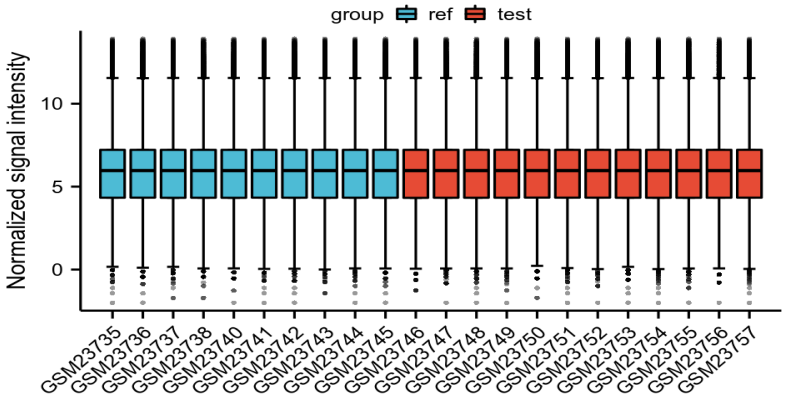


Fig 1. box diagram of GSE 1428 data set

(Note: the first 10 samples are marked in blue, and the last 12 samples are marked in red. The expression of all samples is consistent, and different colors indicate that they are from different samples.)

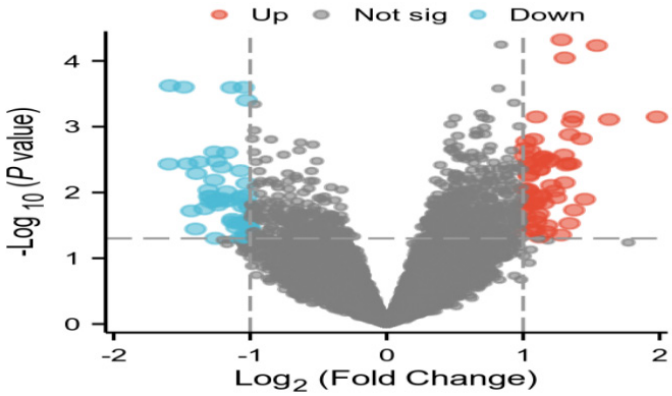


Fig 2. Volcano map

3.2. Enrichment Analysis of GO and KEGG

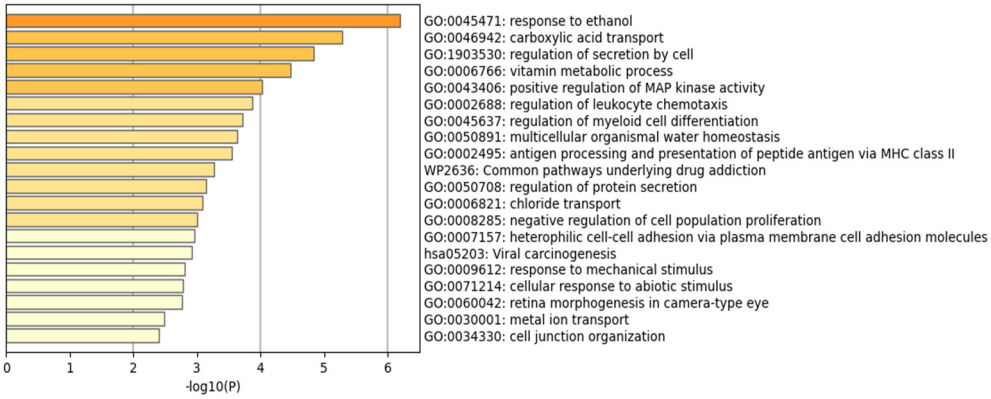


Fig 3. thermogram of GO and KEGG functional enrichment analysis

In order to understand the biological functions of DEGs, this study used Metascape analysis tool to analyze 117 DEGs, including biological processes, cellular components, molecular functions and signal pathway in KEGG. The results are as follows: 1) The enrichment of differential gene functions in biological process, molecular function and cell components: reaction to ethanol, carboxylic acid transport, cell secretion regulation, vitamin metabolism process,

positive regulation of MAP kinase activity, leukocyte chemotaxis regulation and bone marrow cell differentiation regulation; 2) KEGG pathway is mainly enriched in viral carcinogenesis pathway (hsa05203), as shown in Figure 3.

3.3. Construction and Module Analysis of Protein Interaction Network of DEGs

DEGs is analyzed by STRING tool, and the minimum interaction score is set to 0.8, and the relationship diagram consisting of 108 nodes and 44 lines is obtained (as shown in Figure 4). The first four core genes, namely RPE65, TOPORS, IMPG2 and IMPDH1, were calculated by MCC algorithm of Cytoscape 3.9.1 software, in which the expression of RPE65, TOPORS and IMPDH1 core genes was down-regulated in satellite cells of skeletal muscle tissue of SA patients, while the expression of IMPG2 was up-regulated. Use MCODE plug-in to analyze PPI network, as shown in Figure 5.

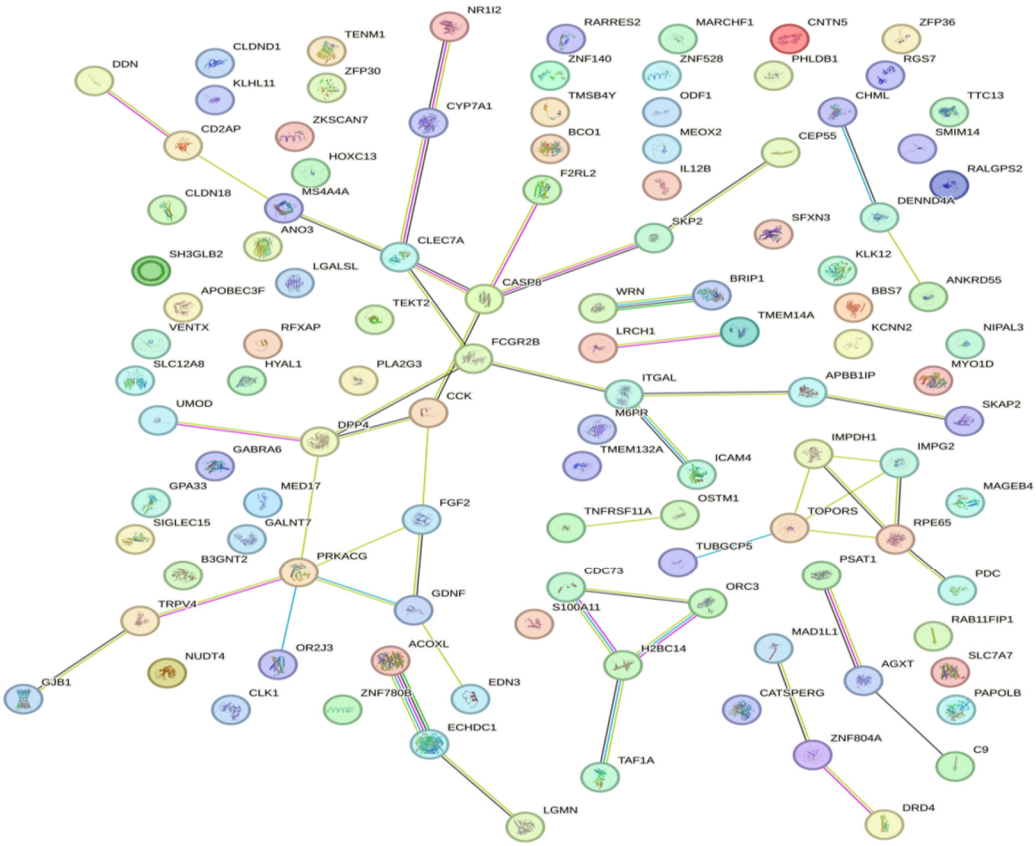


Fig 4. protein interaction network of differential genes.

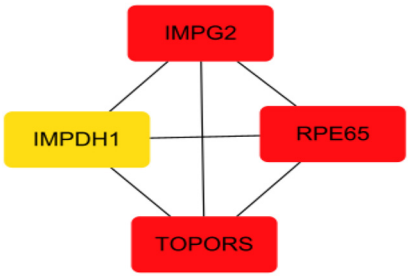


Fig 5. Top 4 Core Genes

3.4. Gene Tissue Localization Analysis

Introducing the core gene into the BioGPS database, it is significant that the highest expression level between tissues is twice higher than the second expression level, and a tissue-specific core gene, RPE65, is screened out, as shown in Figure 6.

4. Conclusion

Muscle mass and strength are heritable, with the heritability of muscle strength ranging from 30% to 85% and muscle mass ranging from 45% to 90%, suggesting that the onset of SA has a certain genetic tendency, which may be related to the expression of specific genes. Therefore, it is particularly important to expand the research scope of SA genomics.

In this study, 22 samples of human lateral thigh muscle in GEO database were analyzed, and four DEGs, including RPE65, TOPORS, IMPG2 and IMPDH1, were obtained. It was found that they were mainly enriched in cell secretion regulation, vitamin metabolism, leukocyte chemotaxis regulation, bone marrow cell differentiation regulation and viral carcinogenesis pathway. It is found that patients with SA often show high levels of white blood cells, and the increase of white blood cells is independently related to the increase of SA, This may be related to the regulation of DEGs on leukocyte chemotaxis. Under the chemotaxis of IL-33, bone marrow cells changed from pro-inflammatory phenotype to pro-regenerative phenotype, regulating the activity of muscle stem cells. Skeletal muscle development needs vitamin D, which is beneficial to maintain muscle strength and reduce the risk of falling. Vitamin D level is related to muscle mass. To sum up, leukocyte chemotaxis, vitamin D metabolism and bone marrow cell differentiation regulation are closely related to the pathogenesis of SA, which coincides with the discovery that core genes are enriched in cell secretion regulation, vitamin metabolism process, leukocyte chemotaxis regulation and bone marrow cell differentiation regulation, which

provides a certain basis for core genes to participate in the pathogenesis of SA.

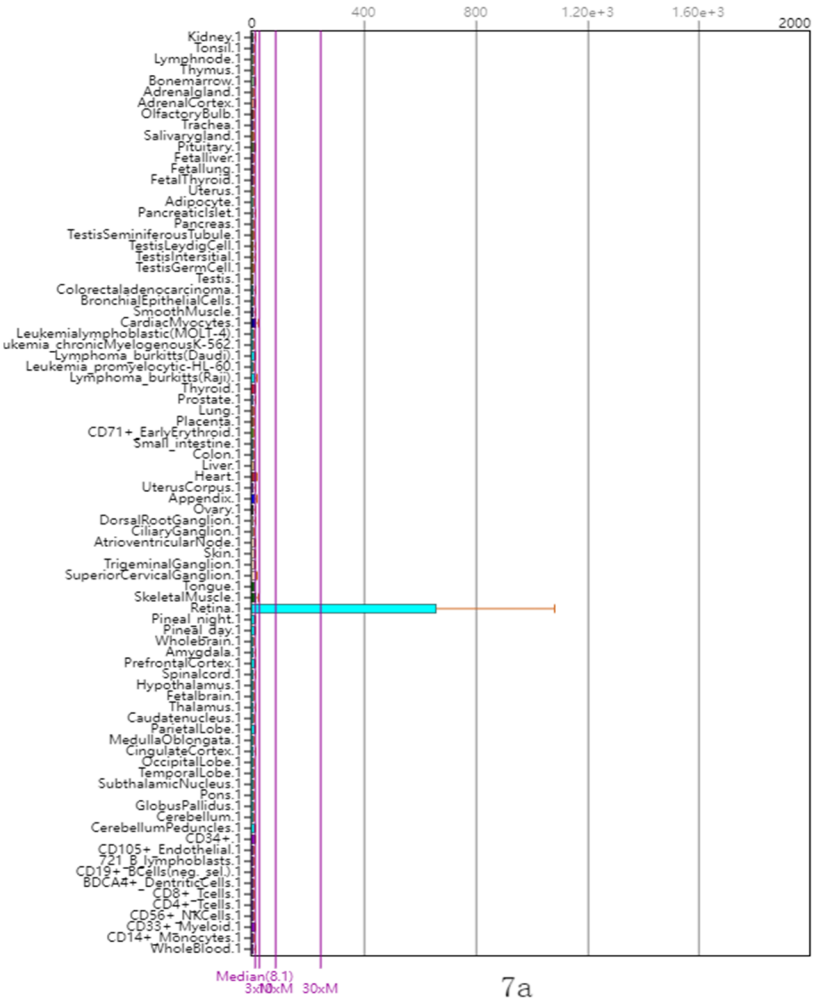


Fig 6. tissue localization analysis of core genes

RPE65 encodes a specific protein of retinal pigment epithelium, which is easily affected by oxidative stress, resulting in RPE65 easily splitting into fragments of different sizes in the synthesis of retinal pigment epithelium, and RPE45 is one of them. Oxidative stress-induced cleavage of RPE65 is regulated by signal transduction pathway mediated by caspase and amyloid precursor protein, which is involved in neurodegenerative diseases and inflammation. It has been proved that muscle apraxia and oxidative stress caused by nerve injury are the main pathogenesis of SA. To sum up, it is speculated that RPE65 is decomposed into proteins of different sizes under the action of external factors, which is involved in regulating the signal pathway mediated by caspase and amyloid precursor protein, causing nerve injury and oxidative stress, leading to SA, which needs further study and confirmation.

TOPORS encode a ring finger protein which is bound by p53 and topoisomerase I, where p53 is a tumor suppressor. The pathogenesis of SA involves different pathways of apoptosis, such as autophagy, calcium-dependent protease and proteasome system increasing protein degradation and decreasing muscle stem cell activation, These pathways are related to the activity of p53, which plays an important role in regulating the differentiation and maturation of muscle cells. P53 is also involved in NF-κB and PI3 K/Akt pathways, which play a central role in regulating myoprotein synthesis and degradation. To sum up, TOPORS has a great influence on TP53 and its encoded p53 protein, and p53 is widely involved in the pathogenesis of SA. Therefore, this study speculated that TOPORS may be involved in the pathogenesis of SA through the regulation of p53, which can be regarded as a direction of future research on SA.

IMPG2 encodes proteoglycans in retinal photoreceptor matrix, i.e. proteoglycans 1 and 2(IMPG1 and IMPG2), and IMPG1 and IMPG2 share the same domain, i.e. SEA domain. SEA has proteolytic activity and is attached to many proteins, including dystrophin (DG) and mucin. DG translation and modification are regulated by SEA, and the defect of DG modification is related to muscular dystrophy, DG is an important component of dystrophin, which stabilizes the muscle membrane by acting as an axis connecting extracellular matrix with cytoskeleton. When the glycosylation of DG is destroyed, it will affect the ligand binding, thus leading to progressive muscle degeneration. IMPG2 imbalance may be related to SA, which deserves further study.

Key enzymes in purine nucleotide biosynthesis encoded by IMPDH1, IMPDH1 is expressed in almost all tissues and participates in maintaining the balance of purine pool. It has been found that IMPDH1 is highly expressed in many tumors, especially in advanced tumors, and the prognosis of tumor patients with high expression of IMPDH1 is often relatively poor, It may be because IMPDH1 has a complex effect on tumor immune microenvironment, which affects tumor immune microenvironment stability through m-6-A modified enzyme protein and participates in regulating NF-κB and

other related pathways. Overexpression of IMPDH1 will lead to a decline in the efficacy of immunotherapy. Immune disorder and increased oxidative stress are risk factors for SA and cachexia in patients with advanced cancer. To sum up, IMPDH1 may aggravate oxidative stress by destroying immune homeostasis, but whether it is related to the pathogenesis of SA needs further discussion.

To sum up, the core genes analyzed in this study may affect skeletal muscle through different pathways, which will help us understand the potential causes of SA, identify the low-level skeletal muscle quality and carry out targeted intervention as soon as possible, and provide theoretical basis for studying the pathogenesis and treatment of SA. However, the conclusions of this study still need to be further verified by basic experiments.

References

[1] Kim YC, Ki SW, Kim H, Kang S, Kim H, Go GW. Recent Advances in Nutraceuticals for the Treatment of Sarcopenic Obesity. Nutrients. 2023;15(17):3854.

[2] CAO L, MORLEY J E. Sarcopenia Is Recognized as an Independent Condition by an International Classification of Disease, Tenth Revision, Clinical Modification (ICD-10-CM) Code[J]. J Am Med Dir Assoc, 2016,17(8): 675-677.

[3] HAN P, KANG L, GUO Q, et al. Prevalence and Factors Associated With Sarcopenia in Suburb-dwelling Older Chinese Using the Asian Working Group for Sarcopenia Definition[J]. J Gerontol A Biol Sci Med Sci, 2016,71(4): 529-535.

[4] LIU P, HAO Q, HAI S, et al. Sarcopenia as a predictor of all-cause mortality among community-dwelling older people: A systematic review and meta-analysis[J]. Maturitas, 2017,103: 16-22.

[5] KIRK B, AL S A, DUQUE G. Osteosarcopenia: A case of geroscience[J]. Aging Med (Milton), 2019,2(3): 147-156.

[6] CRUZ-JENTOFT A J, SAYER A A. Sarcopenia[J]. Lancet, 2019,393(10191): 2636-2646.

[7] GOLJANEK-WHYSALL K, IWANEJKO L A, VASILAKI A, et al. Ageing in relation to skeletal muscle dysfunction: redox homoeostasis to regulation of gene expression[J]. Mamm Genome, 2016,27(7-8): 341-357.

[8] PELLEGRINO A, TIIDUS P M, VANDENBOOM R. Mechanisms of Estrogen Influence on Skeletal Muscle: Mass, Regeneration, and Mitochondrial Function[J]. Sports Med, 2022,52(12): 2853-2869.

[9] PAN L, XIE W, FU X, et al. Inflammation and sarcopenia: A focus on circulating inflammatory cytokines[J]. Exp Gerontol, 2021,154: 111544.

[10] Chechenova M, Stratton H, Kiani K, et al. Quantitative model of aging-related muscle degeneration: a Drosophila study. Preprint. bioRxiv. 2023;2023.02.19.529145.

[11] CHUNG T H, SHIM J Y, LEE Y J. Association between leukocyte count and sarcopenia in postmenopausal women: The Korean National Health and Nutrition Examination Survey [J]. Maturitas, 2016,84: 89-93.

[12] Im YG, Han MY, Baek HS. Association of Serum Vitamin D Level with Temporomandibular Disorder Incidence: A Retrospective, Multi-Center Cohort Study Using Six Hospital Databases. Nutrients. 2023;15(13):2860.

[13] Papadopoulou SK. Sarcopenia: A Contemporary Health Problem among Older Adult Populations. Nutrients. 2020;12 (5): 1293.

[14] Conte E, Dinoi G, Imbrici P, De Luca A, Liantonio A. Sarcoplasmic Reticulum Ca2+ Buffer Proteins: A Focus on the Yet-To-Be-Explored Role of Sarcalumenin in Skeletal Muscle Health and Disease. Cells. 2023;12(5):715.

[15] Li Q, Ishii KA, Kamoshita K, et al. PAC1 Deficiency Protects Obese Male Mice From Immobilization-Induced Muscle Atrophy by Suppressing FoxO-Atrogene Axis. Endocrinology. 2023;164(6): bqad065.

[16] PEI J, GRISHIN N V. Expansion of divergent SEA domains in cell surface proteins and nucleoporin 54[J]. Protein Sci, 2017,26(3): 617-630.

[17] CHOI J Y, LOPES L, BEN M C, et al. Maturation of the malarial phosphatidylserine decarboxylase is mediated by high affinity binding to anionic phospholipids[J]. J Biol Chem, 2023, 299(5): 104659.

[18] KEPPEKE G D, ANDRADE L, BARCELOS D, et al. IMPDH-Based Cytoophidium Structures as Potential Theranostics in Cancer[J]. Mol Ther, 2020,28(7): 1557-1558.

[19] RUAN H, SONG Z, CAO Q, et al. IMPDH1/YB-1 Positive Feedback Loop Assembles Cytoophidia and Represents a Therapeutic Target in Metastatic Tumors[J]. Mol Ther, 2020,28(5): 1299-1313.

[20] LIU C, ZHANG W, ZHOU X, et al. IMPDH1, a prognostic biomarker and immunotherapy target that correlates with tumor immune microenvironment in pan-cancer and hepatocellular carcinoma[J]. Front Immunol, 2022,13: 983490.